AD-A264 295

REPORT DOCUMENTATION PAGE

Form Approved OMB No 0704-0188

(2)

...on is estimated to average 1 hour per response including the time for reviewing instructions searching existing data sources ...leting and reviewing the collection of information Send comments regarding this burden estimate or any other aspect of this ...ducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson ...and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED FINAL 1 Feb 89 - 28 Feb 92

4. TITLE AND SUBTITLE

"RESEARCH ON COLLABORATIVE PLANNING" (U)

5. FUNDING NUMBERS

61102F

2304/A7

6. AUTHOR(S)

Dr. Barbara J. Grosz

DTIC ELECTE MAY 14 1993 S C D

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Harvard University
20 Aiken Computation Lab
Cambridge MA 02138

8. PERFORMING ORGANIZATION REPORT NUMBER

AFOSR-TR- 93 0211

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFOSR/NM
110 Duncan Ave Suite 100
Bolling AFB DC 20332-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AFOSR-89-0273

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release;
Distribution unlimited

12b. DISTRIBUTION CODE

UL

13. ABSTRACT (Maximum 200 words)

This project investigated characteristics of collaborative problem-solving by multiple human agents, determined the properties needed by a computer system to participate in collaborative plan-based activities, and designed formalizations for representing and reasoning about multi-agent actions and collaborative plans. They defined a set of core action relations, designed an action representation language and a representation that provides for incrementally building action representations from partial information, and significantly modified the SharedPlan model of collaborative activity (GS90) to provide for a greater variety of action relations and more complex act-types.

93 5 18 043

93-10746

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

RESEARCH ON COLLABORATIVE PLANNING

Final Report to the Air Force Office of Scientific Research
Contract number: AFOSR-89-0273

February 1, 1989 through February 29, 1992

Barbara J. Grosz, Principal Investigator

Project Participants:
Cecile Balkanski
Joyce B. Friedman
Karen Lochbaum
Candace L. Sidner

Abstract

This project investigated characteristics of collaborative problem-solving by multiple human agents, determined the properties needed by a computer system to participate in collaborative plan-based activities, and designed formalizations for representing and reasoning about multi-agent actions and collaborative plans. We defined a set of core action relations, designed an action representation language and a representation that provides for incrementally building action representations from partial information, and significantly modified the SharedPlan model of collaborative activity [GS90] to provide for a greater variety of action relations and more complex act-types. Copies of technical papers reporting work supported by this project are included with this report. We have also appended to this report descriptions of our work on mutual beliefs, negotiation in collaborative activity, and modelling of intentions all of which have not yet been published. The report itself summarizes our results.

1 Summary of Results

Our milestones for this project were to collect interaction records of planning by two agents, to analyze the actions and action relations in the data, to identify a test domain for implementation of a collaborative planning system, to formulate definitions of a core set of actions and action relationships in the test domain, to determine representations of belief and intention for use in modelling collaborative activity using SharedPlans, to specify conditions under which negotiation plays a role in establishing intermediate goals for collaborative planning, to investigate the use of constraint-based formalisms for reasoning about action relations, to develop algorithms that used an act-type lattice in incrementally constructing action descriptions, and to develop a model of intentions that was useful for reasoning about the intentions of collaborating agents.

We analyzed data from the following three sources: a construction task, a group planning meeting, and a simulated human-computer problem-solving dialogue. An analysis of the actions and action relations in this data was performed, and an action representation language was developed [Bal90]. We chose network management as our main test domain, but the implementations discussed below incorporated representations of actions from domains in the AI planning literature (e.g.[Kau90]) as well as those in the test domain of network management.

We developed an intermediate ("logical form" like) representation that allows for incrementally building action representations from partial information and designed procedures for determining the action relations implicit in sentences about multiple actions [Bal91, Bal92a, Bal92b]. We began implementation of a system: this implementation is being completed under funding obtained after this project was completed. It will be reported in an upcoming Ph.D. dissertation [Bal93].

We extended and significantly modified the SharedPlan model of collaborative activity described in [GS90] to provide a basis for the construction of computer systems that can work with people in jointly solving tasks that require coordinating the combined capabilities of systems and users. The assumptions underlying this model are less restrictive along a number of dimensions than those of other approaches to multi-agent planning and plan recognition. In particular, we assume that different agents have different knowledge and different capabilities, whereas most other approaches assume identical capabilities and many others assume identical knowledge of actions and the ways in which they are performed; we assume that agents share control of the problem-solving process and no one agent is in charge throughout a task, whereas many approaches assume either a managing agent or complete autonomy; and we assume that the agents are communicating and collaborating on deciding the way in which to achieve their objectives and on who should perform different subtasks, whereas other approaches assume no collaboration on the problem-solving process

itself.

The revised definition of SharedPlans provides for action relations other than generation, and enables handling of act-types that require multiple kinds of action relationships in the specification of the ways in which they can be accomplished. Two constructs play central roles in this revision: *recipes* and the general act-type relation, *Contributes*. To model collaborative planning also requires that we specify algorithms that model the process of an agent augmenting the beliefs and intentions in a partially specified collaborative plan. This process comprises the adoption of mutual beliefs about actions in the recipe for the plan, properties of those actions, and intentions to perform them, as well as the actual adoption of these intentions. This revised definition was used in a companion project supported by U S WEST Advanced Technologies for which we designed and implemented a SharedPlan augmentation algorithm that differs significantly from previous work on plan recognition; it utilized ideas from constraint-based formalisms. This work is described in several conference papers [LGS90, Loc91].

Finally, we investigated a variety of formalizations of mutual belief and intention. A brief summary of our findings on mutual belief is included in Appendix A. Our investigation of representations of intentions suggested the need for substantial additional research, which we could only begin under this project; the issues raised are summarized in Appendix C.

A Mutual Belief in SharedPlans

We have been investigating a variety of formalizations of mutual belief to determine the one most appropriate for use in the context of our use of the construct of SharedPlans to model collaborative planning. Although we had originally intended to devise a representation of intentions as well, we discovered several generalizations and extensions were needed to model mutual belief adequately for SharedPlans. In addition, we determined that it would be useful to extend the definition of SharedPlans from two agents to an arbitrary (finite) set of agents, and hence we needed to develop a formalization of the mutual beliefs of a set of agents. In the remainder of this section we describe our work on that problem.

We began with the system B introduced by Pollack for her study of simple plans[Pol90] and then studied a series of systems to a determine a modal system for mutual belief. The properties of this modal system were studied.

The initial question we asked was how best to extend the system B with the basic predicate BEL to a system for mutual belief for use in collaborative planning. The definition of having a SharedPlan is written using a mutual belief predicate, MBEL(G1,G2,p), where G1 and G2 are two individual agents and p is a proposition.

However, SharedPlan is intended to be symmetric in the agents, so it seems reasonable to move to the notation MBEL({G1,G2},p), displaying the symmetry. But then it is natural to extend to an arbitrary (finite) set of agents, giving MBEL(G,p). We then added to the axioms of B a (rather standard) iterative definition of mutual belief, as well as a rule of induction and the necessary arithmetic to allow us to reason with the definition. In this new system B+ we have proved analogues of most of the axioms of B in terms of MBEL. The analogue of the negative introspection axiom does not hold for MBEL, the other analogues do hold. However, the Barcan formula was needed in proving the analogue of the positive introspection axiom. So, since positive introspection is wanted for the intended application, the Barcan formula was added as an additional axiom. One can then prove all of the theorems that seem to be needed for working with mutual belief in SharedPlans. Belief itself is just the mutual belief of a singleton set. If the notion of SharedPlan is later extended to collaborative plans of more than two agents, it will be useful to be able to prove theorems about subsets of the initial set of agents, as is possible in B+.

As Barwise [Bar88] observes, most accounts of mutual belief are "after characterizations of ... mutual belief in terms of belief". However, in the work on SharedPlans, mutual belief is inferred in discourse by conversational postulates [GS90], not by proofs from belief. Typically, agent x asserts a proposition p, agent y agrees (or fails to disagree within a reasonable time) and mutual belief of x,y in p is then inferred. There is no intermediate reasoning in terms of belief. This suggested that one might move to a system in which mutual belief is taken as basic, and belief is later defined. In our next system, MB, we attempt to carry out this proposal. Axioms were given in terms of MBEL, and a definition of belief derived. G is the agent of the predicate MBEL (rather than a set of agents of BEL) and partial order of set inclusion is used.

For the system MB there are a few immediate metatheorems: MB is an extension of B and MB is consistent relative to B+. The important metatheorem of completeness seems hard to obtain. We do not see how to prove completeness with respect to B+, since the definitions of mutual belief in B+ involve terms not available in MB. In fact, as of now we do not believe that MB is complete, and we seek an augmentation of the axiom set that will make it complete. Another approach is to try to show completeness with respect to a possible worlds semantics.

To do this, we rewrite the system MB as the modal system M, making explicit the modal operators that were treated in MB as though they were predicates. The system M involves a set of modal operators M_G, with a partial order defined on the agents G. The agents at the bottom of the partial order correspond to what previously was thought of as singleton sets.

B Negotiation in Collaborative Activity

We characterized the type of negotiation relevant to SharedPlans, which we call collaborative negotitation. Collaborative negotiations occur between two agents who have established a shared high-level goal and must engage in some activities, including verbal actions, to achieve this goal. Collaborative negotiations are distinguished from negotiations in which agents have high-level conflicting goals, as, for example, in labor disputes [Syc87]. We have identified several sub-types of collaborative negotiations that may be distinguished by the type of beliefs relevant to collaborative planning that the participants must hold.

To provide an initial description of collaborative negotiations, we have identified those portions of sample dialogues in the network management domain in which agents negotiate about the actions to be undertaken or the objects involved in these actions. (These objects correspond to grounding the parameters of act-types in the recipes of the SharedPlan.) These dialogues exhibit one type of collaborative negotiation, that in which the agents mutually know several recipes for the goal, but disagree either about the appropriate choice of recipe or about the objects to be used in the particular recipe they've chosen. We also identified data for further study for a second type of collaborative negotiation in which the agents do not already agree on the recipe and hence must negotiate at the very highest level about the way in which to accomplish their task.[1]

C Intention

Our research into the types of actions that collaborating agents perform [Bal90] and the kinds of intentions needed to collaborate revealed several limitations in the definition of SharedPlans (repeated in Figure 1). Some of these limitations became evident only when we attempted to extend our algorithms from simple actions to complex actions with several levels of decomposition. Others were made evident through our analysis of models of intention and a determination of the need to model the commitment-to-act property of intention. These problems resulted from SharedPlans originally being designed only for recipes in which each constituent act is performed by one of two agents, and used in a discourse context in which an agreement to cooperate was used to establish the SharedPlan.

As a result of the second assumption, an assumption of the circumstance of use of SharedPlans, the current definition of having a SharedPlan may be satisfied by the beliefs and intentions of non-cooperative agents. The problem may be seen by con-

[1] Subsequent to leaving this project, Sidner has continued this research at the DEC Cambridge Research Laboratory [Sid92].

SharedPlan(G_1,G_2,A,T1,T2) $\Longleftrightarrow$

1. MB(G_1,G_2,R:instantiation-of-recipe-for-**A**,T1)
2. MB(G_1,G_2,EXEC(α_j,G_{α_j},T_{α_j}),T1)
3. MB(G_1,G_2,INT(G_{α_j},α_j,T_{α_j},T1),T1)
4. MB(G_1,G_2,INT(G_{α_j},α_j $\wedge$ Contributes(α_j,**A**),T_{α_j},T1),T1)
5. INT(G_{α_j},α_j,T_{α_j},T1)
6. INT(G_{α_j},α_j $\wedge$ Contributes(α_j,**A**),T_{α_j},T1)

Figure 1: The definition of SharedPlan

sidering an example that appears in a paper of Searle's on "collective intentionality," that appeared concurrently with our original presentation of SharedPlans [Sea90]. The example concerns a group of MBA students each of whom intends to help humanity by pursuing his or her own selfish interests. In the first version of this example (Version A), the students form a pact on graduation day to help humanity in this way. In the second version (Version B), the students simply mutually believe that they each have the same intention; they have not formed any agreement to do so or to work together for a common purpose. Searle argues that Version A is a case of collective intentionality, whereas Version B is not. In Version A, each student pursues his or her own selfish interest as a means of *their collectively* helping humanity. In the second variation, however, each student's effort is a means of *his* or *her individually* helping humanity.

Hobbs [Hob90] argues that the SharedPlan definition incorrectly models Version B by attributing collaboration where there is none: although Hobbs's argument itself appears to be flawed, further investigation has shown his conjecture that SharedPlans may incorrectly ascribe collaboration where there is none to be true. The Contributes relation in Clauses (4) and (6) of the SharedPlan definition was intended to model the collective intention that holds only of Version A, e.g. *I* intend to perform an action, α_j, so as to contribute to *our* performing **A**. However, the formalization of the relation shown in the definition does not include a representation of agents; as a result, the representation also models Version B. That is, because the Contributes relation is specified over act-types, this representation does not actually require that the act-type **A** be performed by both agents. Although the Contributes relation could simply be modified to hold of activities (which include an agent representation) rather than act-types, blindly doing so introduces the undesirable situation whereby one agent

intends for another agent to do a particular act, something which we argued against in our original formulation [GS90].

The correct specification of the Contributes relation is actually only part of a larger set of issues: the formalization of a *group's intention* to perform an action. The action representation used in the SharedPlan model [Bal90] allows for the representation of multi-agent activities (as is necessary in modelling collaborative activity), but the SharedPlan definition itself only treats recipes in which each constituent act is performed by one of two agents.

To develop a formal model of intention adequate to cover joint activity, it is necessary to determine what that model should include and how its contents should differ from the contents of the SharedPlan definition. For example, suppose that four agents (g1, g2, g3, and g4) are collaborating to build a porch swing in such a way that two of the agents, g1 and g2, are going to put together the base of the swing and the other two, g3 and g4, are going to put together the back. To model the collaboration of the four agents, the SharedPlan should include a mutual belief that agents g1 and g2 intend to build the base and that agents g3 and g4 intend to build the back; however, it should not include mutual beliefs of any agent's individual intentions. That this is correct may be seen, for example, from the fact that g1 does not need to know what part of building the back g3 is going to do, but only needs to believe that the building of the back will be performed. The intentions of individual agents do appear in subsidiary SharedPlans that are undertaken to satisfy jointly held intentions; for instance, g3 and g4's SharedPlan to build the back would include each of their individual intentions to perform parts of that building event as well as mutual belief of those intentions.

The SharedPlan definition presumes a model of intention in which an agent can intend only his own actions [GS90]. Although the current definition is sufficient when all constituent actions of a recipe can be performed by a single agent, it does not extend easily to more complex constituent acts. Approaches to modelling joint intention have disagreed on the relationship that may hold between the set of agents intending an action and the set of agents performing that action. For example, Cohen & Levesque [CL91] allow for a set of agents to intend that one of its members perform an action, while Searle [Sea90, Hob90] argues for a collective intention in which an individual can intend that a group of which that individual is a member perform an action. Hobbs [Hob90] suggests that the "allowable" relationships between the sets of agents depends on whether the intentions are future-directed intentions or intentions-in-action [Bra90]. There is as yet no formalization that is adequate for the kinds of complex activities we are investigating.

Initial research into the issues discussed above has also illustrated that, in its formal definition, the SharedPlan model of collaboration lacks explicit reference to the cooperativeness of the agents. This cooperativeness is instead modelled in several

conversational default rules [GS90]. These rules are used in establishing the mutual beliefs necessary for two agents to agree to undertake a SharedPlan. Although cooperativeness is implied by the action representation used in SharedPlans (e.g. the activity ⟨lift(piano), {joe,pam},t1⟩ represents that Joe and Pam lift the piano together at time t1), further research includes a more careful inspection of the degree of cooperativeness implied by recipes and the possible need to incorporate other aspects of cooperativeness into the SharedPlan definition itself.

Thus, SharedPlans combine three separable aspects of collaborative planning and activity: the need to model complex actions in which more than one agent may be involved (and hence recipes that provide for multiple agents of constituent acts), the need to coordinate intentions of multiple agents, and the need to explicitly model the commitment to act. The definition of SharedPlans must be revised to distinguish more clearly among these different components. We expect the recipe clause [Clause (1)] to become more precise in the process of this reformulation.

References

[Bal90] Cecile T. Balkanski. Modelling act-type relations in collaborative activity. Technical Report TR-23-90, Harvard University, 1990.

[Bal91] C.T. Balkanski. Logical form of complex sentences in task-oriented dialogues. In *Proceedings of the 29th Annual Meeting of the ACL, Student Session*, Berkeley, CA, 1991.

[Bal92a] Cecile T. Balkanski. Actions, beliefs and intentions in rationale clauses and means clauses. In *Proc. of AAAI92*, 1992. To appear.

[Bal92b] Cecile T. Balkanski. Actions, beliefs and intentions in rationale clauses and means clauses. In *Proceedings of AAAI-92*, pages 296–301, 1992.

[Bal93] Cecile Balkanski. *Actions, Beliefs, and Intentions in Rationale Clauses and Means Clauses*. PhD thesis, Harvard University, 1993. (to appear).

[Bar88] J. Barwise. Three views of common knowledge. In *Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge*. Morgan Kaufmann, 1988.

[Bra90] Michael E. Bratman. What is intention? In *Intentions in Communication*, chapter 2. The MIT Press, 1990.

[CL91] P. Cohen and H. Leveque. Teamwork. Technical Report 503, SRI International, 1991.

[GS90] B.J. Grosz and C.L. Sidner. Plans for discourse. In P.R. Cohen, J.L. Morgan, and M.E. Pollack, editors, *Intentions in Communication*. Bradford Books at MIT Press, 1990.

[Hob90] J. R. Hobbs. Artificial intelligence and collective intentionality: Comments on Searle and on Grosz and Sidner. In P.R. Cohen, J.L. Morgan, and M.E. Pollack, editors, *Intentions in Communication*. MIT Press, 1990.

[Kau90] Henry Kautz. A circumscriptive theory of plan recognition. In P.R. Cohen, J.L. Morgan, and M.E. Pollack, editors, *Intentions in Communication*. MIT Press, 1990.

[LGS90] K.E. Lochbaum, B.J. Grosz, and C.L. Sidner. Models of plans to support communication: An initial report. In *Proceedings of the Eighth National Conference on Artificial Intelligence - AAAI-90*, Boston, MA, 1990.

[Loc91] K. E. Lochbaum. An algorithm for plan recognition in collaborative discourse. In *Proceedings of the 29th Annual Meeting of the ACL*, Berkeley, CA, 1991.

[Pol90] Martha E. Pollack. Plans as complex mental attitudes. In P.N. Cohen, J.L. Morgan, and M.E. Pollack, editors, *Intentions in Communication*. Bradford Books, MIT Press, 1990.

[Sea90] J.R. Searle. Collective intentionality. In Cohen, Morgan, and Pollack, editors, *Intentions in Communication*. MIT Press, Cambridge, MA, 1990.

[Sid92] C. L. Sidner. Using discourse to negotiate in collaborative activity: An artificial language. 1992. Workshop on Cooperation among Heterogeneous Agents, NCAI-92.

[Syc87] K.P. Sycara. *Resolving adversarial conflicts: An approach integrating case-based and analytical methods*. PhD thesis, Georgia Institute of Technology, School of Information and Computer Science, 1987. Also appeared as Tech.Report GIT-ICS-87/26.